Justification Bias in Self-Reported Disability: New Evidence from Panel Data

Nicole Black[a], David W. Johnston[a], Agne Suziedelyte[b, *]

[a] Centre for Health Economics, Monash University, Australia

[b] Department of Economics, City, University of London, UK

Abstract

The relationship between health and work is frequently investigated using self-assessments of disability from social surveys. The complication is that respondents may overstate their level of disability to justify non-employment and welfare receipt. This study provides new evidence on the existence and magnitude of justification bias by exploiting a novel feature of a large longitudinal survey: each wave respondents are asked identical disability questions twice; near the beginning and end of the face-to-face interview. Prior to answering the second disability question, respondents are asked a series of questions that increase the salience of their employment and welfare circumstances. Justification bias is identified by comparing the variation between the two measures within-individuals over time, with the variation in employment status over time. Results indicate substantial and statistically significant justification bias; especially for men and women who receive disability pensions.

Keywords: justification bias; disability; non-employment; panel data

---

This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Social Services (DSS) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper, however, are those of the authors and should not be attributed to either DSS or the Melbourne Institute.

* Corresponding author: agne.suziedelyte@city.ac.uk@city.ac.uk; +44 (0)20 7040 4543; Department of Economics, City, University of London.

1. Introduction

Understanding the relationship between health and work is central to labor and health economics research and crucial for the design of health policies, social welfare systems, and strategies for productivity and growth. This relationship is often investigated using self-assessments of health and disability from social surveys. However, there exists a legitimate concern that thresholds for reporting a work-limiting disability may vary systematically according to individual circumstances (Kapteyn et al. 2007). In particular, individuals without a paid job may overstate their health-related work limitations because of financial incentives, such as qualifying for a disability pension. It is also possible that social context and psychological factors compel the non-employed to use illness to rationalize their inability to fulfil a socially prescribed role (Shuval et al. 1973). This so called 'justification bias' implies that the estimated importance of health and disability on labor supply decisions is most likely inflated. To more precisely measure the role of health in economic decision making, it is therefore critical to ascertain the magnitude of justification bias and characterize the types of individuals for whom justification bias is largest. In this paper, we present new evidence on these issues.

Despite the long-running recognition and attention devoted to the issue of justification bias, there is conflicting evidence about its importance. In early investigations, Anderson and Burkhauser (1985, p.324) state "we are persuaded that self-reports of health are unsatisfactory measures", while on the other hand, Stern (1989, p.392) concludes that "standard disability measures are powerful and reasonably exogenous predictors of labor force participation". A decade later, Kerkhofs and Lindeboom (1995) and Kreider (1999) find substantial over-reporting of work limitations, whereas, Dwyer and Mitchell (1999) find no evidence in support of the justification hypothesis using an overall general self-assessed health indicator and only weak evidence of justification bias using self-reported work limitations. More recently, Benítez-Silva et al. (2004, p.649) are "unable to reject the hypothesis that self-reported disability is an unbiased indicator", while in contrast, Baker et al. (2004, p.1090) find "evidence that the error in self-reported chronic conditions is related to labor market status", and the results in Lindeboom and Kerkhofs (2009, p.1042) "show that justification bias is substantial and that failing to account for this may change estimation results considerably". Further recent evidence on the importance of justification bias can be found in Gannon (2009), Datta Gupta and Larsen (2010), Datta Gupta and Jürges (2012), and Gosling and Saloniki (2014).

We contribute to this diverse literature by using an approach that differs from previous studies. We exploit a unique feature of an Australian longitudinal survey in which disability status is self-reported twice in each wave using identical questions – once at the beginning and once at the end of the face-to-face interview.[1] This question identifies disabilities or health conditions that have lasted six months or more, restrict everyday activity, and cannot be corrected by medication. The second disability question is, however, preceded by a series of questions about employment and sources of income, including disability welfare. Therefore, it is likely that respondents are inadvertently 'primed' to consider these issues when reporting disability the second time. How survey design can induce or exacerbate misreporting of health and disability has received little acknowledgement in the justification bias literature, but it has been shown that responses to life evaluation questions are extremely sensitive to question-order effects (Deaton, 2012). Priming has also been used in economic experiments to increase the salience of certain concepts and issues (Benjamin et al., 2010; Callen et al., 2014; Cohn et al., 2015).

[1] Previous studies have exploited repeated health questions in surveys to investigate reporting bias and heterogeneity; see Crossley and Kennedy (2002), Clarke and Ryan (2006) and Lumsdaine and Exterkate (2013). In these studies the questions regard general health (rather than disability status) and either the survey mode (i.e. face-to-face versus self-completion), question wording, or available response options differ between the two survey questions. In addition, none of the studies use longitudinal data.

The second novel feature of our approach is that we fully utilize the panel dimension of our data by estimating fixed-effects (FE) regression models. Essentially, we investigate how within-individual changes in the variation between the two self-reported disability measures correlate with within-individual changes in employment status. This modelling approach allows us to control for all time-invariant factors that influence reporting behavior, such as survey design and cognitive ability.

The FE results demonstrate that non-employed respondents and disability pension recipients are significantly more likely to exaggerate their level of disability. For example, we find that conditional on the response to the disability question at the beginning of the interview, unemployed and out of labor force (OLF) males are 3.1 and 6.6 percentage points respectively more likely to report a disability at the end of the interview than are employed males. The corresponding effects are smaller for females (2.2 and 2.6 percentage points respectively). For men and women, effect sizes are larger for respondents receiving disability pension payments (including employed pension recipients), but are still substantial and statistically significant for unemployed and OLF respondents not receiving pensions.

2. Causes of Justification Bias

Justification bias is a form of state-dependent reporting, whereby the reporting of self-assessed disability (or health) is systematically related to one's employment status. Most commonly, justification bias is the tendency for non-employed individuals to over-report their disability level, relative to their true or latent disability, in order to rationalize their economic inactivity. One potential motivation for this behavior is financial. For example, respondents who are fraudulently collecting disability-related welfare payments may overstate their disability. This inflation of self-reported disability may be motivated by a fear that their survey responses could be used by officials to re-assess their welfare eligibility.[2] Moreover, even if respondents understand that social surveys are not designed to assess or monitor welfare eligibility, they may feel a social desire to justify their welfare receipt to the interviewer.

Another possible cause of justification bias is the desire to conform to socially accepted norms associated with different states of employment (Myers 1982). This desire induces inadvertent subtle changes of thresholds for equating poor health with a disability. For example, an employed respondent suffering migraines may not usually consider themselves as having a work-limiting disability, but after they become non-employed, their threshold for what they consider a work-limiting disability decreases and their assessment changes.[3] Another less conventional example is when an employed respondent under-reports their true disability. The social norm that workers are physically robust and capable of performing their paid roles can lead employed respondents to increase their disability reporting threshold.

Related to both the financial and social causes is the desire by respondents to present themselves in the best possible light during interviews. This drives respondents to exaggerate socially desirable behaviors or characteristics and underreport those that are less desirable. This is known as social desirability bias (Bowling 2005). In our context, non-employed respondents (regardless of whether they receive a disability pension or not) may feel that a disability (or ill health) is a more socially acceptable reason for non-employment than either their failure to find employment or their choice to not work. Therefore, whether they are unemployed, early retired or out of labor force for other reasons, respondents who feel a social obligation to be working

---

[2] Parsons (1982, p.83) observed that "The self-rated poor health group will be composed of two distinct subsets: those who would rate themselves in poor health in an incentive-neutral environment, and those who are induced by the economic environment to declare themselves in poor health."

[3] This mechanism is based partially on the concept that "disability" is not an objective binary health state, but more so a categorization that is based on self, doctor, or government evaluations and definitions. As Autor and Duggan (2006; p.85) write: "While certain medical conditions are clearly disabling, "disability" is not a medical condition. Disability is a dividing line (or zone) chosen by policymakers on a continuum of ailments affecting claimants' capability to engage in paying work and their pain and discomfort in doing so", and "Beyond the subset of clearly incapacitating medical and mental disorders, the extent of "disability" is ultimately a variable determined by policy."

may inflate their level of disability. We would expect that this behavior is more likely to occur if the respondent's employment status is at the forefront of their mind or if they are conscious of the interviewer knowing their employment status.

The use of illness to legitimize one's failure to fulfil a socially prescribed role has been recognized for some time (Shuval et al. 1973). However, we still know very little about how the social pressure to justify non-employment varies across individuals, and in turn how this may lead to heterogeneity in justification bias. Given the traditional gender roles around providing income for the family, we may expect greater social pressure on males to use poor health as a reason for not working. Indeed males have been the sole focus in many studies that examine justification bias (e.g. Lindeboom and Kerkhofs 2009). While few studies have explicitly examined gender differences, there is some evidence to suggest that over-reporting of disability among non-workers (aged 50-64) is higher for women than men (Kreider 1999). This suggests that other, less obvious, social pressures may also be playing a role, and that further investigation into gender differences is important.

We may also expect to see differences in the tendency for justification bias by age, ethnicity, and education level. For example, non-employment may be more socially acceptable among older individuals near retirement age, than among younger individuals in the prime of their working lives. Cultural norms about contributing to household income and accepting welfare may differ, and therefore we may expect heterogeneity in justification bias by ethnicity or country of birth. Social class, often measured by education level, may also influence the social pressure felt by individuals to justify their non-employment.

Another factor that may influence justification bias is the rate of unemployment in an individual's area. There is some evidence that self-rated general health of the long-term unemployed is better in high unemployment regions, and worse in regions with low rates of unemployment (Whitehead et al. 2005). Therefore, in areas where unemployment is high, and consequently more commonplace to be non-employed, individuals may feel less social pressure to justify their inactivity.

3. Methods and Data

3.1. Empirically Identifying Justification Bias

Our aim is to estimate the likelihood that two individuals in different employment states, but with identical 'true' health, report different disability status. This estimate would be easily obtained if individuals' 'true' health was observable; however, to the best of our knowledge, no such data exists.[4] One alternative second-best approach is to conduct a randomized experiment. In such an experiment the treated sample would be 'primed' to consider their employment status prior to answering the disability question. Through such a process, the treated individuals' affiliation with their employment-related identity is increased (made more salient), causing their behavior to shift towards that identity's norms (Benjamin et al., 2016). Importantly, the employment priming effects are interesting not only because of their direct influence on disability reporting behavior, but also because they are informative about the norms associated with different employment states, such as a disability norm for out-of-the-labor-force (OLF) men (Benjamin et al., 2016).[5]

Our approach is similar to the imagined random experiment. Within the longitudinal HILDA survey, which is described in Section 3.2, each year respondents are asked an identical disability question twice by the same interviewer, under the same conditions, within

[4] A close approximate to 'true' health could be obtained through a comprehensive medical examination of each surveyed individual at the time of the survey (potentially including a full physical and mental examination, blood sample, urine sample, and x-rays). Unfortunately, administrative medical records would typically be insufficient for measuring 'true' health, due to the possibility of undiagnosed health problems, and diagnoses that occurred in time periods or in jurisdictions not included in the records.

[5] A work-related priming example is Cohn et al. (2014), who increase the salience of bank employees' professional identity by asking seven questions about their professional background, such as "At which bank are you presently employed?" and "What is your function at this bank?". Treated employees behaved more dishonestly in a coin tossing task than control employees. See Cohn and Maréchal (2016) for a review of the literature that uses priming to study economic questions.

approximately 30 minutes. In between the two questions, respondents are asked questions about their employment status and income sources. We interpret these intervening questions as a form of priming, as they likely increase the salience of respondents' employment and welfare states, and cause respondents to respond to subsequent questions in a manner consistent with the social norms of their employment status. Given this priming, we expect that non-employed respondents are more likely than employed respondents to change their answer and report that they do have a disability when asked the second time. We interpret such behavior as evidence of justification bias.[6]

Importantly, this approach requires two key assumptions: (i) respondents' interpretation and understanding of each disability question is identical, and (ii) respondents' information regarding their own health when answering each question is identical. This latter assumption would be violated if, for example, the respondent had initially forgotten about their disability, but had remembered it by the second disability question, perhaps because of intervening questions. We are confident that these two assumptions hold for three reasons. First, the two questions are identical, and for both questions, the survey respondent is handed the same 'Showcard' to read that explains the criteria for having a disability, and a long list of potential causes, such as "hearing problems", "chronic or recurring pain", and "any condition that restricts physical activity or physical work (e.g. back problems, migraines)" (Appendix A includes a copy of the Showcard). Presumably the list of various health conditions and disabilities are helpful in triggering a respondent's memory. Second, the same interviewer within the same interview (the same sitting) verbally asks both questions. For this reason we

[6] There may be many respondents who are significantly influenced by the survey priming who do not change their disability self-assessment, because their self-assessed health status does not cross their (new) disability threshold. If our measure was instead ordinal (as is sometimes used within the literature), we may have observed greater changes in self-assessed disability. For example, Lindeboom and Kerkhofs (2009) examine the disability question, 'Does your health limit you in the kind and the amount of work that you can do?', which allows four possible responses: causes no problems, causes some problems, causes severe difficulties; and makes it impossible to work. Using this example, survey priming may induce changes from 'causes no problems' to 'causes some problems', however, it is unlikely that the analysis of our binary measure will identify such changes.

can rule-out possible violations due to survey mode effects, which have been shown to impact upon response consistency (e.g. Crossley and Kennedy, 2002). Third, the second question is asked approximately 30 minutes after the first, which means it is implausible that a change in respondents' true health status or new information received about their health status could be driving our results. Note also that there are no health-related survey questions between the two disability questions.

An additional threat to the validity of our approach is the possibility that the intervening survey questions alter respondent reporting behavior in ways that are unrelated to a change in the salience of current employment and welfare states. For example, answering numerous survey questions may reduce the respondent's short-term cognitive resources, or alternatively the respondent may become more comfortable in revealing sensitive information. Fortunately, we are able to control for these alternative 'treatments' using the longitudinal aspect of our data. Specifically, we estimate individual fixed-effects models, in which the non-employment effects on disability reporting are identified from those respondents who have had a change in employment status. Because the survey design stays constant across survey waves, the alternative survey 'treatments' are effectively differenced out. Results from falsification tests support this approach. For example, if respondents become more comfortable in revealing their disability status as the interview progresses, we would expect respondents to increasingly report both work-limiting and non-work-limiting disabilities. However, we find statistically insignificant, near-zero estimated effects for non-work-limiting disabilities.[7]

It is important to also note that null estimated effects should not be interpreted as evidence against the existence of justification bias. There will be no difference between unprimed (question one) and primed (question two) reporting behavior if the respondent always

[7] Interview conditions can vary across waves, however, results from individual fixed-effects models that additionally include interviewer-level fixed-effects and interview-timing fixed-effects (e.g. interview month, day of the week and time of day) are very similar to our main results.

feels the need to justify their employment status. The likely presence of these 'consistent justifiers' within our sample will therefore lead to an underestimate of the extent of justification bias. Fortunately, we are able to investigate this issue by comparing self-reported disability status with partner-reported disability status in a sub-sample of respondents whose disability status is initially reported by their partner. Cohabitating partners and married spouses (which we collectively term 'partners') are ordinarily aware of serious, long-term, untreatable conditions suffered by their partners, but are less likely to feel the need to justify their partners' non-employment than their partners themselves. The limitation of using partner reports is that they are less likely to satisfy the assumptions discussed above. That is, it is more likely that the interpretation and understanding of the disability question will differ between partners.

3.2. The Household, Income and Labour Dynamics in Australia (HILDA) Survey

We use data from 12 waves of the HILDA Survey, an ongoing nationally-representative longitudinal study of Australian households that began in 2001. Wave 1 contained a sample of 19,914 panel members from 7,682 households, and in each following year members of these households have been followed-up, along with new household members resulting from changes in the composition of the original household and new households from the wave 11 top-up sample. Data from each year includes detailed information on income, employment, health and other demographic and socio-economic information.[8] The survey comprises three face-to-face survey instruments – Household Form, Household Questionnaire, and Person Questionnaire – and a confidential self-completion questionnaire.

Disability status is first elicited in the Household Form, which we label throughout as Questionnaire 1 (Q1). Q1 is an initial face-to-face questionnaire designed to record basic

[8] The household response rates range from 87.0 per cent in wave 2 to 70.8 per cent in wave 11, while the household response rates for those households responding in the previous wave ranges from 87.0 per cent in wave 2 to 96.4 per cent in wave 11 (Summerfield et al., 2012).

information about each member of the household before commencing the detailed questionnaires. It is administered to one adult member of the household, which can vary from wave-to-wave, and takes on- average 6 minutes to complete. Among partnered respondents, Q1 is more often completed by female partners (around 60% of cases). Most significantly, the Q1 respondent is asked the following question: “does anyone here have any long-term health condition, disability or impairment such as these?”, and is shown a card with the description “Disabilities/health conditions which have lasted, or are likely to last, 6 months or more; restrict everyday activity; and cannot be corrected by medication or medical aids” followed by a list of 17 types of disability.[9] The Q1 respondent answers “yes” or “no” for all household members, starting with him- or her-self.

Disability status is subsequently elicited in the Person Questionnaire, which we label throughout as Questionnaire 2 (Q2). Q2 is the main survey instrument and is administered face-to-face to every member of the household aged 15 years and over. It takes on-average 33 minutes to complete and contains sections on family background, education, employment, income, family formation, and health, with the health section occurring near the end of the questionnaire. At the very beginning of the health section, respondents are presented with a disability card identical to that used in Q1 (same description and list of disability types), and are asked whether they have any long-term health condition, impairment, or disability. Again, they can only answer “yes” or “no”. If the respondent answers “yes”, they are then asked which of the 17 disabilities they have (multiple types can be provided) and whether the condition

[9] The 17 disability types are: Sight problems not corrected by glasses or contact lenses; Hearing problems; Speech problems; Blackouts, fits or loss of consciousness; Difficulty learning or understanding things; Limited use of arms or fingers; Difficulty gripping things; Limited use of feet or legs; A nervous or emotional condition which requires treatment; Any condition that restricts physical activity or physical work (e.g. back problems, migraines); Any disfigurement or deformity; Any mental illness which requires help or supervision; Shortness of breath or difficulty breathing; Chronic or recurring pain; Long-term effects as a result of a head injury, stroke or other brain damage; A long-term condition or ailment which is still restrictive even though it is being treated or medication is being taken for it; Any other long-term condition such as arthritis, asthma, heart disease, Alzheimer’s disease, dementia etc.

limits the "type of work or the amount of work you can do?" These follow-up questions were not asked in Q1.

Within each wave of HILDA, each respondents' disability status is therefore measured twice (in Q1 and Q2). Q1 is either self-reported (64% of cases) or partner-reported (36% of cases), while Q2 is always self-reported.[10] In our analyses, we primarily use the sub-sample of respondents with two self-reports. We are unaware of any comparable data set that repeatedly asks identical disability or health questions using the same survey mode, especially not consistently across waves. HILDA data therefore provide a unique opportunity to investigate reporting heterogeneity in self-reported disability.

3.3. Descriptive Results

Table 1 presents the proportions of working-age individuals (aged 18-60 years) who have a disability according to self-reports in Q1 and Q2. The disability rates are also presented for three employment states: employed (E), unemployed (U) and out-of-the-labor-force (OLF). A respondent is defined as unemployed if they want to work, and are actively looking for work or available to start work within four weeks. The OLF category includes respondents who do not want to work, and respondents who want to work but are not actively looking and are not available to start work. This category includes persons who are retired, homemakers, carers, disabled, travelling / on holiday, and volunteers.[11]

The summary statistics reveal several interesting features of the data. First, self-reported disability rates are around 20 percentage points for both males and females, which is similar to the United States (Kreider and Pepper, 2008). Second, Q1 disability rates are lower than Q2 disability rates. Third, the absolute differences between Q1 and Q2 disability rates are larger

---

[10] Other in-frequent combinations also occur, e.g. a parent may be the Q1 respondent. We do not use these combinations in our analyses.

[11] Given our focus on employment, students are excluded from the sample.

for unemployed and OLF respondents than for employed respondents. For example, the percentage point differences for men equal 1.5, 3.9 and 4.5 for employed, unemployed and OLF, respectively. There is not, however, a similar employment gradient in the proportional differences between the Q1 and Q2 disability rates.

Figure 1 illustrates the extent of variation between the self-reported Q1 and Q2 measures. Specifically, Figure 1A presents the probability of reporting a disability in Q2, conditional on not reporting a disability in Q1: $Pr\left(D_{it}^{Q2} = 1 \middle| D_{it}^{Q1} = 0\right)$. If reporting was consistent across Q1 and Q2, we would expect the rates to be near zero, but instead we observe a remarkable pattern. Figure 1A shows that the propensity for male respondents to change their assessment and report a disability in Q2 increases sharply with non-employment. Conditional on a previous self-report of no disability in Q1, the Q2 disability rates equal 5% for employed, 11% for unemployed, and 21% for OLF. In contrast, the gradient is nearly flat for women. The male pattern suggests that having to rationalize non-employment significantly decreases non-employed men's thresholds for equating poor health with a disability.

Figure 1B presents the opposite variation; the probability of not reporting a disability in Q2, conditional on reporting a disability in Q1: $Pr\left(D_{it}^{Q2} = 0 \middle| D_{it}^{Q1} = 1\right)$. The figure again shows a steep gradient, but in the reverse direction, and for both men and women. Conditional on a previous self-report of an existing disability in Q1, the probability of *not* reporting a disability in Q2 equals 20% for employed men, 7% for unemployed men and 3% for OLF men. The corresponding figures for women equal 17%, 9% and 6%. These patterns suggest that having to detail employment conditions, significantly increases employed respondents' thresholds for equating poor health with a disability.[12]

[12] For completeness, in Appendix Figure B1 we present the graphs that show the conditional probabilities of consistent reporting of disability status (by employment status and gender).

Table 2 presents estimates of individual characteristics that predict variation between the Q1 and Q2 self-reported measures. Using a multinomial logit specification, we model the four mutually exclusive, joint disability outcomes: (1) $D_{it}^{Q1}=0, D_{it}^{Q2}=0$ (no-no); (2) $D_{it}^{Q1}=1, D_{it}^{Q2}=1$ (yes-yes); (3) $D_{it}^{Q1}=0, D_{it}^{Q2}=1$ (no-yes); and (4) $D_{it}^{Q1}=1, D_{it}^{Q2}=0$ (yes-no). Average partial effect estimates are presented for the two inconsistent outcomes (no-yes and yes-no).[13] The results show that non-employed men and women are significantly more likely to answer "no" on Q1 and "yes" in Q2, than their employed counterparts. For men the estimated differences equal 2.1 percentage points for unemployed and 3.4 percentage points for OLF. These effects are large compared to other significant predictors in the model, such as age (0.1 percentage points per year), university degree (-1.9 percentage points), Australian-born (1.1 percentage points), and comprehension issues (2.6 percentage points). The corresponding estimated effects on the converse yes-no outcome are small and statistically insignificant.

Equivalent multinomial logit marginal effect estimates are presented in Appendix Table B2 for the sample of respondents with partner-reported Q1 disability status (Q1 partner-reported, Q2 self-reported). The estimates again show that unemployed and OLF respondents are significantly more likely to have a no-yes outcome than are employed respondents. The effect sizes are slightly larger than in Table 2. For example, the estimated effects for unemployed and OLF men equal 3.5 and 5.9 percentage points, respectively.

The descriptive results presented in this subsection suggest that changes in the likelihood of reporting a disability are associated with employment status. However, these results are unlikely to reflect the true extent of justification bias. As explained in Section 3.1, the questions between Q1 and Q2 may alter how respondents answer Q2 in ways that are unrelated to an increase in the salience of respondents' employment and welfare states. To

[13] Average partial effect estimates for the two consistent outcomes (no-no and yes-yes) are presented in Appendix Table B1.

control for these effects, we rely on within-individual variation over time in the difference between responses to Q1 and Q2.

4. Main Results

As discussed above, our methodological approach involves comparing the variation between self-reported disability status in Q1 (recorded at the beginning of the survey) and Q2 (recorded at the end of the survey) over time, with the variation in employment status over time. We do this by estimating fixed-effects linear probability regression models:

$$D_{it}^{Q2} = \beta D_{it}^{Q1} + E'_{it}\gamma + X'_{it}\delta + \alpha_i + \mu_t + \varepsilon_{it} \tag{1}$$

where D_{it}^{Q2} is disability status from Q2, D_{it}^{Q1} is disability status from Q1, E_{it} is a vector of employment and welfare states, X_{it} is a vector of additional control variables[14], α_i is an individual-level fixed-effect, μ_t is a time fixed-effect, and ε_{it} is a random error term. Statistically significant positive effects of non-employment on self-reported disability are then interpreted as evidence of justification bias.

Estimates of β and γ from equation (1) are reported separately by gender in columns (1) and (3) of Table 3. In Panel (A) the employment vector includes indicators of being unemployed and out of the labor force (OLF). In Panel (B) employment status is interacted with current Disability Support Pension (DSP) receipt, with 'Employed without DSP' used as the omitted category.[15] As background information, in Australia 16-64 year-olds are eligible

[14] The control variable set includes a quadratic function in age, having dependent children, and interview conditions (being suspicious about the study or uncooperative; comprehension issues, such as, poor eyesight or hearing, reading difficulties, or language issues; being interviewed in follow-up fieldwork period; needing more than four calls to complete all interviews; and length of the interview).

[15] Between the Q1 and Q2 disability measures, respondents are asked whether they currently receive the Disability Support Pension (DSP) and how much they received in their most recent payment. We use this self-reported information to construct the DSP indicator.

for the DSP if they have a physical, intellectual, or psychiatric condition that prevents them from working 15 hours or more per week within the next 2 years. In our sample time-frame, eligibility was based on a report from the claimant's doctor. The maximum fortnightly DSP payment for singles without children equals A$782, which is substantially higher than the unemployment benefit (A$519).[16] DSP recipients may work up to 30 hours per week and continue to receive a part pension.

Column (1) of Table 3 shows that conditional on self-reported disability from Q1, unemployed men are 3.1 percentage points and OLF men are 6.6 percentage points more likely to report a disability in Q2 than are employed men. The corresponding estimates for women in Column (3) equal 2.2 percentage points and 2.6 percentage points. The estimates of β, the coefficient on self-reported Q1 disability equal 0.538 to 0.590 for males and females, respectively, which shows that the two disability self-reports are correlated, but not perfectly so.[17] Appendix Table B3 presents the estimated effects of the control variables, most of which are not statistically significant. The probability of reporting a disability in Q2 is higher for respondents who have had comprehension issues during the interview or taken longer to complete Q2. Females with dependent children and older men are also less likely to report a disability in Q2.

The estimates in Panel (B) indicate that the effects are particularly large for respondents receiving the DSP: employed, unemployed and OLF male DSP recipients are 9.3 percentage points, 14.4 percentage points and 14.2 percentage points more likely than employed men not receiving the DSP to report a Q2 disability, conditional on Q1 disability status. The effects are similarly large for female DSP recipients. Importantly, the estimates are also statistically significant for unemployed and OLF men and women who do not receive the DSP. This

[16] For more information see www.humanservices.gov.au/customer/services/centrelink/disability-support-pension and http://www.humanservices.gov.au/customer/services/centrelink/newstart-allowance.
[17] Hausman test confirms that the fixed-effects model is the appropriate estimator. It rejects the equality of random-effects and fixed-effects models for both genders and specifications (p-values < 0.0005).

suggests that social (and not solely financial) incentives triggered by the employment 'priming' questions are likely to play an important role in the inflation of disability status. We can be quite confident that non-employed non-DSP recipients are unlikely to be truly disabled. This is because DSP payments are substantially higher than unemployment benefit (UB) payments, and involve considerably fewer obligations to maintain eligibility (Saunders, 2007). Therefore, truly disabled non-employed individuals have a strong incentive to apply for the DSP.

In columns (2) and (4) of Table 3 we report estimates using an alternative dependent variable. After the main disability question in Q2, respondents are asked whether the condition limits the "type of work or the amount of work you can do?" Respondents who respond 'yes' are considered to have a 'work-limiting disability', and respondents who respond 'no' are considered to have a 'non-work limiting disability'. If respondents are primed by the questions regarding employment and welfare receipt, the estimated effects of non-employment on work limiting conditions should be especially large. Columns (2) and (4) of Table 3 shows that all the 'work-limiting disability' estimates are larger than the corresponding 'any disability' estimates. For example, the estimates for OLF respondents not receiving DSP increase from 6.4 to 8.9 percentage points for men, and from 2.2 to 3.3 percentage points for women. Moreover, the estimated effects are remarkably large for non-employed men receiving DSP. Conditional on self-reported disability from Q1, these men are almost 20 percentage points more likely to report a work-limiting disability in Q2, than are employed men not receiving DSP.

In contrast to the large estimates shown in Columns (2) and (4), the estimated effects of non-employment and DSP receipt on the probability of reporting a 'non-work-limiting condition', compared to no disability, are all small and statistically insignificant. Specifically, the estimated effects for unemployed and OLF men equal -0.004 (t-statistic = -0.46) and -0.012 (t-statistic = -0.73), and the estimated effects for unemployed and OLF women equal -0.001 (t-statistic = -0.19) and <0.0005 (t-statistic = 0.09); these results are not shown in Table 3, but are available upon request. The null results for non-work-limiting conditions are encouraging, because they suggest that alternative explanations for our findings – for example, that respondents become more comfortable in revealing true disabilities as the interview progresses – are unlikely, because if this were the case, we would expect to also observe increased reporting of non-work limiting conditions.

It is also reassuring that the pattern of results shown in Table 3 is duplicated when alternative identification approaches are utilized. In Table 4 we replicate the specifications from Table 3, but instead rely on the sub-sample of respondents whose disability status is initially reported by their partner. As discussed in Section 3.1, cohabitating partners and married spouses are less likely to feel the need to justify their partners' non-employment than their partners themselves, and so therefore, the estimated effects using this sample may be less attenuated by respondents who justify in both Q1 and Q2. Overall, the results in Table 4 are very similar to those in Table 3, with the main exception being that the effect of DSP receipt for women seems less important. For example, the estimated effect for female OLF DSP recipients decreases from 0.103 (in Table 3) to 0.020 (in Table 4). Another alternative modelling approach is to condition the sample on self-reported Q1 disability status rather than to include it as a covariate (as in Table 3). Again, the results from this approach are similar to those shown in Table 3 (see Appendix Table B4). In this case, however, the biggest exception is that the estimated effects of DSP receipt are larger.[18]

5. Exploring Heterogeneity

5.1. Type of Disabling Condition

[18] To further test the robustness of the estimates shown in Table 3, we have estimated fixed-effect models using subsamples of single respondents. For singles, Q1 is necessarily self-reported and therefore the estimated non-employment effects cannot be driven by bias from non-random selection of Q1 respondents. The estimates for singles are very close in magnitude to those from Table 3 (results available upon request).

After the main disability question in Q2, respondents are asked which of the 17 conditions they have. We use answers to this question to construct five categories representing the most frequently chosen, identifiable conditions: (1) mental health conditions; (2) limited use of limbs; (3) condition restricting physical activity; (4) chronic or recurring pain; and (5) hearing and sight problems.[19] Reported conditions not examined here are those that are rarely chosen (e.g. speech problems) and the catch-all options (e.g. any other long-term condition).

Column (1) of Table 5 shows that conditional on self-reported disability from Q1, unemployed men are 1.9 percentage points and OLF men are 7.3 percentage points more likely to report a mental health condition than are employed men. Relative to sample mean levels, these effects are far larger than for any other condition type. One potential explanation for this result is that there is variation across conditions in the effect that priming has on reductions in individuals' disability thresholds. It is probable that workers with poor mental health are less likely to consider themselves disabled than are workers with poor physical health, and are also more likely to continue working (leading to the often discussed issue of presenteeism). Consequently, individuals with poor mental health reduce their disability thresholds to a larger extent when they transition from employment to non-employment, even without a corresponding change in true mental health. The non-employment effects are also statistically significant for 'limited use of limbs', 'condition restricting physical activity' and 'chronic or recurring pain'. The non-employment effects for 'hearing and sight problems' are small and statistically insignificant.

For females the estimated effects are also largest for the 'mental health condition' category: unemployed women are 2.6 percentage points and OLF women are 3.0 percentage

[19] The specific conditions comprising the five categories are: (1) "A nervous or emotional condition which requires treatment" and "Any mental illness which requires help or supervision"; (2) "Limited use of arms or fingers", "Difficulty gripping things" and "limited use of feet or legs"; (3) "Any condition that restricts physical activity or physical work (e.g. back problems, migraines)"; (4) "Chronic or recurring pain"; (5) "Hearing problems" and "Sight problems not corrected by glasses or contact lenses".

points more likely to report a mental health condition than employed women (conditional on Q1 disability).[20] The estimated effects for the categories 'limited use of limbs' and 'condition restricting physical activity' are also large for women, largely mirroring the results for men. Appendix Table B5 shows that disability pension receipt increases the probability of reporting all health conditions for both men and women, but the likelihood of reporting a mental health condition increases most. Appendix Table B6 shows that for men the results remain robust when we use spouse-reports of disability in Q1, but no statistically significant effects are found for women.

The comparatively large effect sizes on the 'mental health condition' category, which is comprised of the two conditions "a nervous or emotional condition which requires treatment" and "any mental illness which requires help or supervision", are particularly interesting given the very large increases over time in the proportions of individuals receiving disability pensions for mental ill-health. The proportion of Australian Disability Support Pension recipients listing a psychological or psychiatric condition as their primary medical condition has risen from 23% in 2001 to 31% in 2013.[21] The growth in disability pension receipt for mental ill-health has also been documented in other developed countries. For example, Autor and Duggan (2006) show that in the US the proportion of Disability Insurance (DI) Awards for the diagnosis group 'mental disorders' has risen from 16% in 1983 to 25% in 2003.

5.2. Individual and local area characteristics

[20] The strong mental health results for both genders are not driven by respondents becoming more comfortable in reporting sensitive information (such as potentially stigmatized mental ill-health) throughout the interview. The estimated effects for work-limiting mental health conditions are even larger, while the estimated effects for non-work limiting mental health conditions are small and statistically insignificant.

[21] The top 5 most commonly claimed for conditions in 2013 are: (1) psychological / psychiatric (31%); (2) musculo-skeletal and connective tissue (26%); (3) intellectual / learning (12%); (4) nervous system (5%); and (5) circulatory system (4%). The distribution of primary medical conditions is similar for both sexes. For more information see the 2013 report by the Australian Government Department of Social Services on "Characteristics of Disability Support Pension Recipients", available at www.dss.gov.au.

As discussed in Section 2, some groups of individuals may be more likely to justify non-employment than others. In particular, social norms, either self-imposed or reflecting the views of society, may play an important role. As may susceptibility to social desirability bias: the desire by respondents to present themselves in the best possible light during interviews. In this subsection, we explore the heterogeneity in justification behavior by age, country of birth, educational attainment, and local area unemployment rate. The results are presented in Table 6. In each regression, we interact the non-employment variables (and Q1 self-reported disability status) with a particular characteristic.[22]

Panel (A) presents results for younger (<40) and older (≥40) respondents. Non-employment is often more socially acceptable among older individuals, who are approaching retirement, than among younger individuals, and so we expect effects to be larger for respondents aged < 40. Overall, there is not strong support for this hypothesis in our data; though the effect for OLF young men is especially high (= 0.094). More generally, these results show that justification behavior is prevalent among individuals of all ages.

Given the potential importance of cultural norms regarding work, we expect to observe some heterogeneity by country of birth. In Panel (B) we split the sample in to those born in non-English speaking (NES) and English speaking (ES) countries. For men, the estimated effects are significantly different between these two groups. The effects for unemployed and OLF men born in English speaking countries are large and statistically significant (0.036 and 0.070, respectively), while the effects for unemployed and OLF men born in non-English speaking countries are smaller and statistically insignificant (-0.001 and 0.037, respectively). It is difficult to isolate the cause of these differences. Interestingly, there exists a stronger belief

[22] We have also tested whether there are significant differences between newer and more experienced respondents, motivated by the possibility that experienced respondents are more likely to trust the HILDA survey, and therefore feel more comfortable revealing sensitive or stigmatized information. Whether using a cutoff of 3 years or 6 years within the HILDA study to define ‘experienced’, we do not find significant differences in effect sizes across experience levels.

among NES men that it is their duty to work: NES men are more likely to rate the following statements as important benefits of paid work: "economic independence (a useful way to serve society)", "not having to be reliant on the Government for income support", and "being able to contribute to the financial costs of maintaining a household." Therefore, it is possible that NES respondents feel a greater need to justify their employment status or welfare receipt whenever asked.

In Panel (C) we analyze the heterogeneity in justification bias by individuals' education (university degree versus lower qualifications). A priori, it is unclear whether education level is correlated with justification bias positively or negatively. Because justification process involves cognitive effort, individuals with a higher cognitive ability and thus educational attainment may be more likely to misreport their disability status. On the other hand, education is correlated with social class and individuals of higher social standing may feel less pressure to justify their non-employment. The results support the latter hypothesis. We find that justification bias is limited to the individuals with lower levels of education. These results are consistent with Kreider (1999) who find substantial over-reporting of work limitations among blue collar workers but no evidence of over-reporting among white collar workers.

In the final panel (D) of Table 6 we explore heterogeneity by the unemployment rate in an individual's local area. The pressure to justify one's unemployment status may be less severe in areas where unemployment is higher and seen as a 'norm', rather than as a personal failure. The results show that there are only small differences across high and low unemployment areas, with inconsistent patterns across genders. The effects are slightly larger for men living in high unemployment areas, and are slightly lower for women living in high unemployment areas.

6. Conclusion

The last two decades has seen the number of disability pension recipients more than double in Australia (Broadway et al. 2014), with similarly worrying trends in the United States (Liebman 2015). Notably, the increase has been especially large for hard-to-verify impairments such as back pain and mental health problems (Liebman 2015). Consequently, the Australian government has recognized that “many new applications for the disability pension are not triggered by the acquisition of an impairment or disability, but by changes in an individual’s employment circumstances” (Macklin 2009). In a similar way, an individual’s employment circumstances can trigger an increase in reported levels of impairment or disability in household surveys. This may be due to financial incentives, but is likely also driven by a social desire to justify non-employment or welfare receipt to the interviewer. The presence of justification bias is problematic in research that relies on self-reported disability (Lindeboom and Kerkhofs 2009) due to the bias it generates in the estimated relationships between health and employment status. Nonetheless, self-reported disability remains an important survey measure of an individual’s capacity to work (Bound 1991).

Our study provides new evidence on the existence and magnitude of justification bias using a novel feature of the HILDA panel dataset that allows us to test whether increasing the salience of an individual’s employment circumstances increases the threshold for equating poor health with a limiting disability. Each wave, respondents are asked an identical disability question twice by the same interviewer, under the same conditions, within approximately 30 minutes. In between the two questions, respondents are asked questions about their employment status and income sources. Through this process, the treated individuals’ affiliation with their employment-related identity is increased (made more salient), causing their behavior to shift towards that identity’s norms, such as a disability norm for out-of-the-labor-force (OLF) men (Benjamin et al., 2016).

We formally identify justification bias using a within-individual fixed effects (FE) approach and control for a range of individual characteristics and interview conditions. This approach, which identifies the effects from respondents who have had a change in employment status, strengthens our ability to isolate justification bias from other sources of reporting heterogeneity. We find that non-employed respondents and disability pension recipients are significantly more likely to misreport or exaggerate their level of disability. For example, we find that conditional on responses to the disability question at the beginning of the interview, unemployed and OLF males are 3.1 and 6.6 percentage points respectively more likely to report a disability at the end of the interview than are employed males. The effects of non-employment on misreporting of disability are generally smaller for females, but still statistically significant. We also find that individuals receiving a disability support pension, including those who are employed, are more likely to exaggerate their disability. For example, employed males receiving a disability pension are 9.3 percentage points more likely to report a disability than are employed males not receiving a disability pension. However, estimated effects are still substantial and statistically significant for unemployed and OLF respondents not receiving pensions.

Self-reported disability will continue to be a practical and informative measure of disability in large surveys; however, we demonstrate that the ordering of questions can have a considerable impact on the reporting of health limitations. Our results show that when a question about having a limiting disability is preceded by questions about employment history, job search, reasons for not working, and pension receipt, individuals have an incentive to change their threshold for equating poor health with a disability, and this incentive differs systematically by employment status. Therefore, to minimize problems associated with justification bias, future surveys should position disability and health questions before questions related to employment and income. Researchers that use surveys where this ordering has not been achieved need to carefully weigh the merits of using disability measures that are affected by justification bias.

References

Anderson, K. H. and R. V. Burkhauser (1985). "The Retirement-Health Nexus: A New Measure of an Old Puzzle." Journal of Human Resources 20(3).

Autor, DH. and MG Duggan (2006). “The growth in the social security disability rolls: a fiscal crisis unfolding.” Journal of Economic Perspectives 20(3): 71-96.

Baker, M., M. Stabile and C. Deri (2004). "What do self-reported, objective, measures of health measure?" Journal of Human Resources 39(4): 1067-1093.

Benítez-Silva, H., M. Buchinsky, H. Man Chan, S. Cheidvasser and J. Rust (2004). "How large is the bias in self-reported disability?" Journal of Applied Econometrics 19(6): 649-670.

Benjamin, D.J., J.J. Choi and G. Fisher (2016). “Religious Identity and Economic Behavior.” The Review of Economics and Statistics 98(4): 316-637.

Benjamin, D.J., J.J. Choi and A.J. Strickland (2010). “Social Identity and Preferences.” American Economic Review 100(4): 1913–28.

Bound, J. (1991). "Self-Reported versus Objective Measures of Health in Retirement Models." Journal of Human Resources 26(1): 106-138.

Bowling, A. (2005). "Mode of questionnaire administration can have serious effects on data quality." Journal of Public Health 27(3): 281-291.

Broadway, B., A. Chigavazira and S. Kassenboehmer (2014). “Labour Force Potential of Disability Support Pension Recipients.” Melbourne Institute of Applied Economic and Social Research Report.

Callen, M, M. Isaqzadeh, J.D. Long and C. Sprenger (2014). “Violence and Risk Preference: Experimental Evidence from Afghanistan.” American Economic Review 104(1): 123–48.

Clarke, P.M. and C. Ryan (2006). “Self-reported health: reliability and consequences for health inequality measurement.” Health Economics 15: 645-652.

Cohn, A., J. Engelmann, E. Fehr, M.A. Maréchal (2015). “Evidence for Countercyclical Risk Aversion: An Experiment with Financial professionals”. American Economic Review, 105 (2): 860-885.

Cohn, A., E. Fehr and M.A. Maréchal (2014). “Business Culture and Dishonesty in the Banking Industry.” Nature 516(7529): 86–89.

Cohn A, M.A. Maréchal (2016). “Priming in economics”. Current Opinion in Psychology 12:17-21.

Crossley, T.F. and S. Kennedy (2002). “The reliability of self-assessed health status.” Journal of Health Economics 21: 643-658.

Datta Gupta, N. and M. Larsen (2010). "The impact of health on individual retirement plans: self-reported versus diagnostic measures." Health Economics 19(7): 792-813.

Datta Gupta, N. and H. Jürges (2012). "Do workers underreport morbidity? The accuracy of self-reports of chronic conditions." Social Science and Medicine 75: 1589-1594.

Deaton, A. (2012). “The financial crisis and the well-being of Americans.” Oxford Economic Papers 64(1): 1-26.

Dwyer, D. S. and O. S. Mitchell (1999). "Health problems as determinants of retirement: Are self-rated measures endogenous?" Journal of Health Economics 18(2): 173-193.

Gannon, B. (2009). "The influence of economic incentives on reported disability status." Health Economics 18(7): 743-759.

Gosling, A. and E. Saloniki (2014). "Correction of misclassification error in disability rates." Health Economics 23(9): 1084–1097.

Kapteyn, A., JP Smith,and A. van Soest (2007). “Vignettes and self-reports of work disability in the United States and the Netherlands.” American Economic Review 97(1): 461-473.

Kerkhofs, M. and M. Lindeboom (1995). "Subjective health measures and state dependent reporting errors." Health Economics 4(3): 221-235.

Kreider, B. (1999). "Latent work disability and reporting bias." Journal of Human Resources: 734-769.

Liebman, J.B. (2015). “Understanding the Increase in Disability Insurance Benefit Receipt in the United States.” Journal of Economic Perspectives 29(2): 123-150.

Lindeboom, M. and M. Kerkhofs (2009). "Health and work of the elderly: subjective health measures, reporting errors and endogeneity in the relationship between health and work." Journal of Applied Econometrics 24(6): 1024-1046.

Lumsdaine, R.L. and A. Exterkate (2013). “How survey design affects self-assessed health responses in the Survey of Health, Ageing, and Retirement in Europe (SHARE).” European Economic Review 63: 299-307.

Macklin, J (2009) “Valuing the future: policymaking for the long term.” Speech for the Per Capita Policy Exchange Conference, Canberra, 21 October 2009.

McGarry, K. (2004). “Health and Retirement: Do Changes in Health Affect Retirement Expectations?” Journal of Human Resources 39: 624-648.

Myers, R. J. (1982). "Why Do People Retire from Work Early." Social Security Bulletin 45(9): 10-14.

Parsons, D. O. (1982). "The Male Labour Force Participation Decision: Health, Reported Health, and Economic Incentives." Economica 49(193): 81-91.

Saunders, P. (2007). “Mutual Obligation, Unemployment and Well-being”. Australian Journal of Labour Economics 10: 167-184.

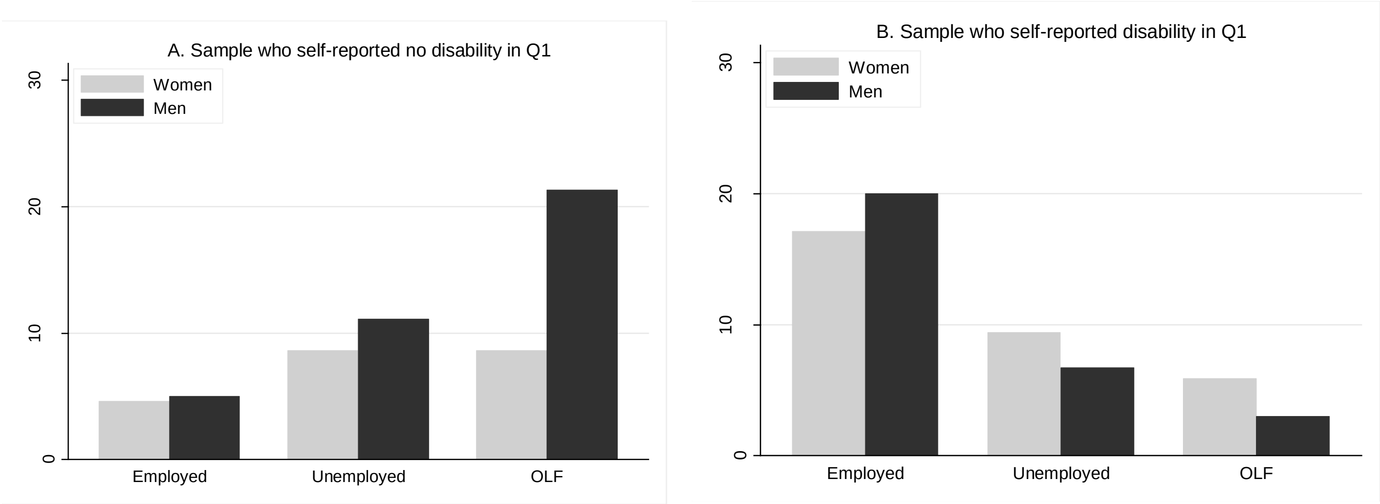
Shuval, J. T., A. Antonovsky and A. M. Davies (1973). "Illness: A mechanism for coping with failure." Social Science & Medicine (1967) 7(4): 259-265.

Stern, S. (1989). "Measuring the Effect of Disability on Labor Force Participation." Journal of Human Resources 24(3): 361-395.

Summerfield M, Freidin S, Hahn M, Ittak P, Li N, Macalalad N, et al. (2012). HILDA User Manual –Release 11. Melbourne: Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Whitehead, M., F. Drever and T. Doran (2005). "Is the health of the long-term unemployed better or worse in high unemployment areas?" Health Statistics Quarterly 25:12.

Figure 1: Proportions of males and females who change their self-reported disability status between questionnaires 1 and 2, by employment status



Note: Samples consist of respondents aged 18-60 for whom we have two self-reported measures of disability.

Table 1: Reported disability rates across questions, genders and employment states

	Male				Female			
	All	E	U	OLF	All	E	U	OLF
Self-report in Q1	0.200	0.141	0.407	0.692	0.195	0.137	0.288	0.368
Self-report in Q2	0.219	0.156	0.446	0.737	0.216	0.153	0.322	0.401
Individuals	4,538	4,179	867	616	6,050	5,117	1,881	2,298
Observations	22,957	19,662	1,618	1,677	35,644	25,333	3,967	6,344

Note: Figures are sample means calculated using our estimation sample of respondents aged 18-60 for whom we have two self-reported measures of disability. The employment status categories are: employed (E), unemployed (U) and out of the labor force (OLF).

Table 2: Average marginal effect estimates from multinomial logit models of disability status recorded in questionnaires 1 and 2

	Males		Females	
	No-Yes	Yes-No	No-Yes	Yes-No
	(1)	(2)	(3)	(4)
Unemployed	0.021***	0.001	0.023***	0.003
	(0.005)	(0.004)	(0.003)	(0.002)
Out of Labor Force	0.034***	0.002	0.019***	-0.001
	(0.005)	(0.005)	(0.003)	(0.002)
Age	0.001***	0.001***	0.001***	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
Has children	0.000	-0.001	-0.003	0.001
	(0.004)	(0.003)	(0.003)	(0.002)
Married/Partnered	0.003	-0.002	-0.000	-0.000
	(0.003)	(0.002)	(0.002)	(0.002)
Born in Australia	0.011***	0.000	0.001	0.003
	(0.004)	(0.003)	(0.003)	(0.002)
University degree	-0.019***	-0.001	-0.007***	-0.004**
	(0.004)	(0.003)	(0.003)	(0.002)
Suspicious/Non-cooperative	0.007	-0.002	0.014**	-0.001
	(0.008)	(0.007)	(0.007)	(0.006)
Comprehension issues	0.026***	-0.001	0.014**	0.003
	(0.006)	(0.006)	(0.006)	(0.005)
Q2 interview length/10	0.007***	0.000	0.008***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Follow up interview	0.006	-0.004	-0.003	-0.002
	(0.005)	(0.004)	(0.004)	(0.003)
More than 4 calls to household	0.001	-0.007***	-0.004	-0.004**
	(0.003)	(0.002)	(0.002)	(0.002)
Observations	22955	22955	35639	35639

Notes: No-Yes = No disability in Q1 but disability in Q2. Yes-No = Disability in Q1 but no disability in Q2. Presented figures are average marginal effects. Standard errors clustered at the individual level are presented in parentheses. Samples consist of individuals for whom disability in Q1 and Q2 are self-reported. All regressions additionally control for year effects. *, ** and *** denote statistical significance at the 0.10, 0.05 and 0.01 levels, respectively.

Table 3: Fixed-effects linear regression models of disability status recorded in questionnaire 2

	Males		Females	
	Any Disability	Work-Limiting	Any Disability	Work-Limiting
	(1)	(2)	(3)	(4)
(A) Unemployed	0.031***	0.045***	0.022***	0.032***
	(0.010)	(0.009)	(0.006)	(0.005)
Out of Labor Force	0.066***	0.090***	0.026***	0.038***
	(0.013)	(0.013)	(0.006)	(0.005)
Self-Reported Q1 Disability	0.538***	0.454***	0.590***	0.531***
	(0.011)	(0.014)	(0.009)	(0.011)
(B) Unemployed - No DSP	0.023**	0.037***	0.020***	0.028***
	(0.010)	(0.010)	(0.006)	(0.006)
Out of Labor Force - No DSP	0.064***	0.089***	0.022***	0.033***
	(0.015)	(0.015)	(0.006)	(0.006)
Employed - Receive DSP	0.093***	0.147***	0.047	0.081***
	(0.023)	(0.026)	(0.030)	(0.030)
Unemployed - Receive DSP	0.144***	0.192***	0.109***	0.154***
	(0.026)	(0.027)	(0.019)	(0.020)
Out of Labor Force - Receive DSP	0.142***	0.194***	0.103***	0.143***
	(0.024)	(0.025)	(0.016)	(0.017)
Self-reported Q1 disability	0.536***	0.449***	0.588***	0.526***
	(0.011)	(0.014)	(0.009)	(0.011)
Observations	22955	21320	35639	33424

Notes: Standard errors clustered at the individual level are presented in parentheses. Samples consist of individuals for whom disability in Q1 and Q2 are self-reported. DSP = Disability Support Pension. All regressions control for a quadratic function in age, marital status, having dependent children, interview conditions (suspicious about the study or uncooperative; problems during interview; interviewed in follow-up fieldwork period; more than four calls to complete all interviews; and length of the interview), and year effects. In Models (1) and (3) the outcome variable equals one if the respondent reports having a disability in Q2 and zero otherwise. In Models (2) and (4) the work limiting disability outcome variable equals one if the respondent reports a work limiting disability in Q2 and zero if the respondent does not have a disability; individuals reporting non-work limiting disability are not included for the estimations of these models. *, ** and *** denote statistical significance at the 0.10, 0.05 and 0.01 levels, respectively.

Table 4: Fixed-effects linear regression models of disability status recorded in questionnaire 2 using the sample of respondents with partner-reported disability status in questionnaire 1

	Males		Females	
	Any Disability	Work-Limiting	Any Disability	Work-Limiting
	(1)	(2)	(3)	(4)
(A) Unemployed	0.033**	0.049***	0.022	0.034**
	(0.014)	(0.013)	(0.015)	(0.013)
Out of Labor Force	0.083***	0.110***	0.031***	0.041***
	(0.018)	(0.018)	(0.011)	(0.010)
Partner-Reported Q1 Disability	0.417***	0.327***	0.438***	0.379***
	(0.013)	(0.015)	(0.017)	(0.019)
(B) Unemployed - No DSP	0.022	0.039***	0.018	0.029**
	(0.015)	(0.014)	(0.015)	(0.013)
Out of Labor Force - No DSP	0.091***	0.119***	0.033***	0.042***
	(0.021)	(0.021)	(0.011)	(0.010)
Employed - Receive DSP	0.060*	0.130***	0.010	0.024
	(0.033)	(0.035)	(0.057)	(0.055)
Unemployed - Receive DSP	0.144***	0.189***	0.078*	0.124**
	(0.037)	(0.039)	(0.046)	(0.048)
Out of Labor Force - Receive DSP	0.118***	0.177***	0.020	0.065*
	(0.026)	(0.029)	(0.038)	(0.038)
Partner-reported Q1 disability	0.416***	0.324***	0.438***	0.378***
	(0.013)	(0.015)	(0.017)	(0.019)
Observations	18872	17379	12035	11247

Notes: Standard errors clustered at the individual level are presented in parentheses. Samples consist of individuals for whom disability in Q1 is partner-reported and Q2 is self-reported. DSP = Disability Support Pension. All regressions control for a quadratic function in age, having dependent children, interview conditions (suspicious about the study or uncooperative; problems during interview; interviewed in follow-up fieldwork period; more than four calls to complete all interviews; and length of the interview), and year effects. In Models (1) and (3) the outcome variable equals one if the respondent reports having a disability in Q2 and zero otherwise. In Models (2) and (4) the work limiting disability outcome variable equals one if the respondent reports a work limiting disability in Q2 and zero if the respondent does not have a disability; individuals reporting non-work limiting disability are not included for the estimations of these models. *, ** and *** denote statistical significance at the 0.10, 0.05 and 0.01 levels, respectively.

Table 5: Fixed-effects linear regression models of specific health conditions reported in questionnaire 2

	Mental Health Condition[a]	Limited Use of Limbs[b]	Condition Restricting Phys. Activity[c]	Chronic or Recurring Pain[d]	Hearing & Sight Problems[e]
	(1)	(2)	(3)	(4)	(5)
(A) Males					
Unemployed	0.019**	0.021***	0.031***	0.020***	-0.000
	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)
Out of Labor Force	0.073***	0.047***	0.059***	0.041***	0.014
	(0.016)	(0.014)	(0.016)	(0.015)	(0.016)
Self-reported Q1 disability	0.218***	0.265***	0.371***	0.269***	0.242***
	(0.018)	(0.019)	(0.018)	(0.019)	(0.018)
Sample mean	0.041	0.051	0.085	0.054	0.046
Observations	15659	15853	16524	15918	15765
(B) Females					
Unemployed	0.026***	0.018***	0.013**	0.009**	0.003
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)
Out of Labor Force	0.030***	0.032***	0.024***	0.016***	0.006
	(0.005)	(0.005)	(0.006)	(0.005)	(0.004)
Self-reported Q1 disability	0.361***	0.318***	0.427***	0.360***	0.222***
	(0.016)	(0.016)	(0.015)	(0.016)	(0.016)
Sample mean	0.051	0.050	0.081	0.057	0.026
Observations	24611	24566	25503	24785	23869

Notes: Standard errors (clustered at individual level) are presented in parentheses. The sample consists of individuals for whom disability in Q1 and Q2 are self-reported. In each model (1) to (5), the outcome variable equals one if the respondent reports having the health condition of interest in Q2 and zero if the respondent reports no disability in Q2; individuals reporting conditions other than the condition of interest are not used for the estimation of the model. All regressions control for a quadratic function in age, marital status, having dependent children, interview conditions (suspicious about the study or uncooperative; problems during interview; interviewed in follow-up fieldwork period; more than four calls to complete all interviews; and length of the interview), and year effects. [a] "A nervous or emotional condition which requires treatment" or "Any mental illness which requires help or supervision". [b] "Limited use of arms or fingers", "Difficulty gripping things" or "limited use of feet or legs." [c] "Any condition that restricts physical activity or physical work (e.g. back problems, migraines)." [d] "Chronic or recurring pain." [e] "Hearing problems" and "Sight problems not corrected by glasses or contact lenses". *, ** and *** denote statistical significance at the 0.10, 0.05 and 0.01 level, respectively.

Table 6: Exploring heterogeneity – fixed-effects linear regression models of disability status recorded in questionnaire 2

	(1) Males		(2) Females	
	Unemployed	OLF	Unemployed	OLF
(A) Age < 40	0.030**	0.094***	0.019**	0.017**
	(0.014)	(0.027)	(0.008)	(0.007)
Age ≥ 40	0.032**	0.058***	0.027***	0.037***
	(0.013)	(0.015)	(0.009)	(0.009)
(B) Born in NES country	-0.001	0.037	0.019	0.029*
	(0.023)	(0.034)	(0.015)	(0.017)
Born in ES country	0.036***	0.070***	0.023***	0.025***
	(0.011)	(0.014)	(0.007)	(0.006)
(C) University degree	0.006	0.007	0.006	0.019
	(0.017)	(0.028)	(0.012)	(0.011)
No university degree	0.038***	0.080***	0.026***	0.028***
	(0.011)	(0.015)	(0.007)	(0.007)
(D) Low unemployment area	0.030**	0.058***	0.032***	0.029***
	(0.014)	(0.016)	(0.009)	(0.007)
High unemployment area	0.032***	0.073***	0.016**	0.023***
	(0.012)	(0.015)	(0.007)	(0.007)
Observations	22957		35644	

Notes: Standard errors clustered at the individual level are presented in parentheses. Samples consist of individuals for whom disability in Q1 and Q2 are self-reported. NES= Non English Speaking; ES= English Speaking. Low unemployment area is defined as below the sample median. All regressions control for self-reported Q1 disability, a quadratic function in age, marital status, having dependent children, interview conditions (suspicious about the study or uncooperative; problems during interview; interviewed in follow-up fieldwork period; more than four calls to complete all interviews; and length of the interview), and year effects. *, ** and *** denote statistical significance at the 0.10, 0.05 and 0.01 levels, respectively.